

**REMARKS**

Applicants have amended claim 16 to correct a typographical error to remove the double period at the end of this claim. Applicants again respectfully request clarification from the Office regarding the rejections set forth in the above-identified Office Action which are primarily based on, "Alshawi (U.S. Patent 581,666)". Applicants believe the Office is referring to U.S. Patent No. 5,815,196 to Alshawi, but respectfully request confirmation from the Office on the identification of the reference the Office is basing its rejections. In view of the above amendments and the following remarks, reconsideration of the outstanding office action is respectfully requested.

The Office has rejected claims 1, 2, 6-10, 14, 15, 17, 18 and 22 - 24 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 581,666 to Alshawi (Alshawi) in view of US 5,818,441 to Throckmorton et al. (Throckmorton), claims 5, 13 and 21 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton and further in view of U.S. Patent No. 6,513,003 to Angell (Angell), claims 3, 11 and 19 under 35 U.S.C. 103(a) as being are rejected under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton and further in view of U.S. Patent 6,647,535 to Bozdagi et al. (Bozdagi), and claim 16 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton and in further view of U.S. Patent 5,900,908 to Kirkland et al. (Kirkland). The Office asserts that Alshawi discloses a method and computer readable medium having stored instructions with at least one processor for providing real-time subtitles [captioning] in an AV signal. The Office asserts that Alshawi includes the automatic conversion of an audio [including speech] signal in the AV signal and that FIG. 1. describes a video-based communications device (5,8). The Office asserts that Alshawi provides segmentation of an AV signal (16) and the further processing of the audio [speech] portion of the signal to provide continuous speech-to-subtitles [speech-to-text] translation (19,21,22) that has the ability to overlay and display text subtitles onto AV signal in real-time [captioning](26). The Office asserts that Alshawi does not disclose synchronizing the caption data with one or more cues in the AV signal, but asserts that Throckmorton teach that a data synchronizing sub-system whose function is to synchronize the primary data stream generated by sub-system 10 with specific associated data. The Office asserts that in Throckmorton input to data synchronizing sub-system 20 is scene information from the primary data stream in the form of timecodes and time durations [cues], and data from associated data generator sub-system 16. The Office asserts that Throckmorton creates a so called script for the delivery and

display of associated data at specific points in time. The Office asserts that it would have been obvious to one of ordinary skill at the time of the invention to modify Alshawi with the synchronization of the caption data with one or more cues in the AV signal as taught by Throckmorton since it would have enhanced the viewing experience of the hearing impaired. The Office also asserts that Alshawi does not show the embedding [encoding] of the text [caption] data within the AV signal, but asserts that Angell teaches the embedding [encoding] of the text [caption] data within the AV signal (Fig. 1(108, 140); Col 4, Line 55 - Col 5, Line 17). The Office also asserts that Alshawi does not show a method of converting the audio portion of the signal to text data that checks whether the amount of caption data is greater than a threshold amount or an expiration time before the process of association occurs, but asserts that Bozdagi et al. show a system and method to enable real-time and near real-time storyboarding on the world wide web. The Office asserts that the combination of Alshawi and Throckmorton do not show portability and the utilization of the device in the classroom, but asserts that Kirkland teaches a method of providing encoding caption data into the program signal.

Alshawi, Throckmorton, Angell, Bozdagi, and Kirkland, alone or in combination, do not disclose or suggest, “converting an audio signal in the AV signal to caption data . . . associating the caption data with the AV signal at a time substantially corresponding to a video signal associated with the converted audio signal in the AV signal, wherein the associating further comprises synchronizing the caption data with one or more cues in the AV signal” as recited in claims 1 and 17 or “a speech-to-text processing system that converts an audio signal in an AV signal to caption data . . . a signal combination processing system that associates the caption data with the AV signal at a time substantially corresponding to a video signal associated with the converted audio signal in the AV signal, wherein the signal combination processing system synchronizes the caption data with one or more cues in the AV signal” as recited in claim 9.

As the Office has acknowledged, Alshawi does not disclose or suggest synchronizing the caption data with one or more cues in the AV signal. In fact, Alshawi at col. 1, lines 39-41 teaches away from the present invention by stating that it is unlikely that any resulting systems will be capable of perfect simultaneous translation and speech synthesis. However, contrary to the Office’s assertions Alshawi in view of Throckmorton does not teach or suggest the claimed invention. The Office’s attention is respectfully directed to col. 3, lines 55-57 in Throckmorton which states, “The term associated data as

used herein refers to a stream of data generated separately from the primary data but having content that is relevant to the primary data in general” (emphasis added). Accordingly, as discussed at col. 4, lines 52-60 in Throckmorton, the data synchronizing sub-system 20 synchronizes the primary data stream generated by subsystem 10 with specific associated data. However, as noted above, the associated data in Throckmorton is separately generated from the primary data stream and thus is not part of the original primary data stream which needs to be synchronized back with the corresponding portion of the primary data stream from which it is derived. There is simply no discussion or suggestion in Throckmorton on how to synchronize data which is converted from an audio signal in the original audio-video clip back with the video signal associated with the converted audio signal. Thus, if as proposed by the Office, Alshawi is considered in view of Throckmorton, at most the cited references would only disclose synchronizing separately generated associated data which does not have any direct correspondence with the primary data stream consisting of the videophone conversation. Similarly, Angell, Bozdagi, and/or Kirkland, alone or in combination with the other cited references, do not teach or suggest the claimed invention.

The problem with the prior art can be illustrated by way of analogy to a movie where the audio portion does not correspond with the movements illustrated in the video portion and thus the resulting audio-visual product is more difficult to watch and comprehend. With captioning, the lack of synchronization between captioning and the corresponding audio in an audio-visual signal makes comprehension of subject matter in the audio-visual signal more difficult, particularly for the hearing impaired. With the present invention, the caption data is synchronized with the cues in the audio-video signal so that the words coming out of the speaker’s mouth on the audio-video signal correspond with the caption data being displayed. As a result, a hearing impaired individual can read the caption data and/or lip read from the speaker’s mouth because the caption data and the audio signal have been synchronized. This synchronization of the caption data with the audio-video signal substantially enhances the ability of a hearing impaired individual to comprehend and retain the content of the audio-video signal. One example of the types of cues which can be added to the video signal to accomplish this is described in paragraph 20 in the above-identified patent application.

Accordingly, in view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 1, 9, and 17. Since claims 2-3 and 5-8 depend from and contain the limitations of claim 1, claims 10-11 and 13-

16 depend from and contain the limitations of claim 9, and claims 18-19 and 21-24 depend from and contain the limitations of claim they are patentable in the same manner as claims 1, 9, and 17.

Alshaw, Throckmorton, Angell, Bozdagi, and Kirkland, alone or in combination, also do not disclose or suggest, “determining a first amount of data in the caption data . . . providing the caption data for the associating when the first amount is greater than a threshold amount or when a first period of time has expired” as recited in claims 3 and 19 or “wherein the speech-to-text processing system further comprises a counter that determines a first amount of data in the caption data and a timer that determines when a first period of time has expired, wherein the speech-to-text processing system providing the caption data for the associating when the first amount is greater than a threshold amount or when timer indicates that the first period of time has expired” as recited in claim 11.

As the Office has acknowledged, Alshaw and Throckmorton do not show a method of converting the audio portion of the signal to text data that checks whether the amount of caption data is greater than a threshold amount or an expiration time before the process of association occurs. However, contrary to the Office’s assertions Alshaw in view of Throckmorton in and further in view of Bozdagi does not teach or suggest the claimed invention. Bozdagi relates to a method and system for automatically parsing a video data signal to identify a subset of representative frames from a set of frames. More specifically, as disclosed at col. 5, lines 11-25 in Bozdagi, an image significance determiner 40 decides whether a selected frame within a segment should be kept as a representative image for that segment. Bozdagi discloses at col. 4, lines 44-67 the use of a frame difference determiner that computes the difference between two consecutive frames on a pixel by pixel basis which is then used to select a representative frame. Additionally, Bozdagi discloses at col. 6, lines 13-35, identifying a frame as a representative frame if the change in intensity between consecutive frames is determined to be greater than a threshold. Further, Bozdagi discloses at col. 7, lines 33-35, an extended time lapse between command data can also trigger the image significance determiner 40 determining that an additional representative image is required. Accordingly, Bozdagi discloses a parsing method and system for selecting a subset of representative of frames based on a comparison of adjacent frames, a comparison of an intensity of a frame against a threshold, or an expiration of a time period. However, these disclosures in Bozdagi have nothing to do with determining when there is a sufficient amount of caption data to begin associating the caption data with the audio-video signal. There is

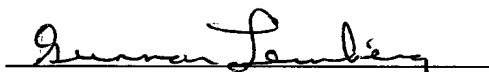
simply nothing in Bozdagi related to controlling the timing of the association of the caption data with the audio video signal. Additionally, Bozdagi is focused on image signals, not audio or caption data from the audio. If, as proposed by the Office, Alshawi is considered in view of Throckmorton and further in view of Bozdagi, at most the cited references would only disclose parsing the videophone conversation so that only certain frames of the videophone conversation were actually transmitted. Similarly, neither Angell nor Kirkland, alone or in combination with the other cited references, teach or suggest the claimed invention.

With the present invention, the rate at which the caption data is associated with the audio video signal is controlled to provide a quick and automated captioning of an audio and visual signal. One example of this process is illustrated in FIGS. 2, 4, and 5 and is described in paragraphs 28 and 32-37 in the above-identified patent application. Accordingly, in view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 3, 11, and 19.

In view of all of the foregoing, applicant submits that this case is in condition for allowance and such allowance is earnestly solicited.

Respectfully submitted,

Date: January 24, 2005

  
Gunnar G. Leinberg  
Registration No. 35,584

NIXON PEABODY LLP  
Clinton Square, P.O. Box 31051  
Rochester, New York 14603-1051  
Telephone: (585) 263-1014  
Facsimile: (585) 263-1600